GeneNetwork/WebQTL Workshop
Sunday, November 12, 2006
20th International Mouse Genome Conference
Chapel Hill, NC, USA

Elissa J Chesler
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Kenneth F Manly
University of Tennessee Health Science Center, Memphis, TN, USA

Self-paced introduction to GeneNetwork/WebQTL

Finding variation in gene expression

Gene transcription is determined by genetic, environmental and gene x environment interactions such as age and sex. Genetic variation in gene expression can be dramatic. The GeneNetwork provides several large data sets of gene expression estimates, most of them derived from recombinant inbred mouse lines. Numerous genes have expression differences between strains ranging over 4 fold.

Standard GeneNetwork links

Point your browser to www.genenetwork.org. This brings you by default to the Find Records page, from which you can retrieve data from many GeneNetwork data sets.

Before you click on any of the record, review the banner of terms near the top of most GeneNetwork pages (Home, Search, Help, News, References, Policies, Accounts, Links). Each of these headings has a pop-down menu from which you can select additional resources and tools. For example, the Search menu heading lists the Search Databases or Find Records page (our starting point), the SNP Browser tool, the Interval Analyst table, and GenomeGraph display. The Policy menu tells you who generated the data sets in GeneNetwork and explains how to contact the data providers and how to use and cite data. The Help menu includes additional tutorials and a set of answers to frequently asked questions. See the last section of this document for more detail on these menu items.

Search and data retrieval

We will focus on the data set defined by: Species: *Mouse*, Group: *BXD*, Type: *Whole Brain mRNA*, Database: *INIA Brain mRNA M430 (Jan06) PDNN*

Enter "Kcnj*" into the ALL or ANY field and click the *Search* button. Note the location and annotation of available potassium channel genes in the Search Results page that opens.

Notice the action of the *Select All*, *Select Invert*, and Select *None* buttons on this page. Select several of the potassium channel entries by clicking the checkbox at the left of the probeset description and/or using those buttons..

BXD Trait Collection page

Click the *Add to Collection* button. The selected probesets will appear in a new window, the BXD Trait Collection page. This is one of two pages that serve as control centers for a variety of analysis

functions. The analysis functions controlled from this page are designed for a set of related traits. These multiple-trait analysis tools are invoked by the buttons at the bottom of the list, buttons marked *Multiple Mapping*, *Compare Correlates*, *Correlation Matrix*, *QTL Cluster Map*, and *Network Graph*. We will explore these later.

Notice that the transcript descriptions on the BXD Trait Collection page include the name of the data set from which they came. This page will accept traits from any BXD data set and from more than one data set. Gene expression traits from different tissues can be combined with classical traits on this page and compared or analyzed together.

For now, close the BXD Trait Collection page by clicking the close button. The information on the BXD Trait Collection page is not lost when the page closes; the page can be re-opened with the selections by choosing *Search* > *Trait Collections* > *BXD Collection* from the menu on any GeneNetwork page.

Use the browser back button to return to previous page, enter "Grin2b" into ALL or ANY field and again click *Search.* In the new Search Results page, note multiple probe sets related to this gene.

Three of the search results are sets of probes that were specifically designed by Affymetrix to estimate the expression of particular pieces of the *Grin2b* mRNA. The probe sets target the "last exon", "intron 2 or a rare exon", and the "far 3' UTR". This information regarding the precise target regions (exon, intron, 3' UTR, etc.) has been collected by the GeneNetwork team. We have only checked the "targets" for a comparatively small number of probe sets. This information that follows the ";" character is valuable because it tells you that the probes target the correct gene. When these data are missing, you should verify location yourself. GeneNetwork contains a simple tool that you will encounter later (Verify by UCSC) that makes it easy for you to check or double-check the genetic and molecular targets of probes. This is an important part of the quality control process mentioned previously. The fourth probe set for *Kif17* is included in this Search Results page because its gene description includes the parenthetical note "NR2B/GRIN2B NMDA receptor transporter."

Trait Data and Analysis Form

Click anywhere on the term *ProbeSet/1422223_at* (this text will highlight as you mouse over it). This is the probe set that targets the last exon of the *Grin2b* transcript. Your click will bring up the Trait Data and Analysis Form. This is the other page that serves as a control panel for a variety of analysis functions.

Notice that this page has three major sections—an upper section with trait information and links, a middle section with analysis tools, and a lower section with phenotypic values for the trait you have chosen.

Characterizing variation in genes and their expression

Gene information links

Examine the upper part of the Trait Data and Analysis Form. For gene expression traits, this section provide basic information about the gene represented by the probe set and it provides links to information at other sites. Explore the links to Entrez Gene, OMIM, UniGene, and Entrez Nucleotide through the links marked *Gene*, *OMIM*, *UniGene*, and *Genbank*, respectively.

This part of the page also has links to other databases that provide information on gene function, location or relationships. Two of these databases are especially useful for neurobiology. Links to these databases take you directly to information about the gene described on this page.

SymAtlas: This database, developed by the Genomics Institute of the Novartis Research Foundation, provides annotation of gene functions and data supporting that annotation.

STRING: is the Search Tool for the Retrieval of Interacting Genes/Proteins, developed and hosted by the European Molecular Biology Laboratory.

PANTHER: This classification system (Protein ANalysis THrough Evolutionary Relationships) is a database resource that classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence.

SynDB: This is a database for the molecular biology of the synapse proteome. It contains a comprehensive collection of proteins that are known or predicted to be associated with synaptic activities.

ABA: The Allen Brain Atlas is an interactive, genome-wide image database defining locations of gene expression in the mouse brain developed by the Allen Institute for Brain Science.

In addition, eight buttons appear in this part of the page:

SNP Browser: This link provides you with a list of all known SNPs in *Grin2b* that are in the GeneNetwork database. You can, of course, also search for SNPs in other genes or regions. The GeneNetwork SNP database is well populated (about 6 million SNPs) and includes all of the Celera SNPs and many SNPs from Perlegen and NIEHS. As of May 2006, there are seven SNPs in *Grin2b* that distinguish C57BL/6J from DBA/2J. Click on the button to have a quick look in a separate window.

GeneWiki: This link lets you annotate our databases. You can leave yourself notes and comments about particular genes or probe sets. You can easily find your own notes using a special search string described in the Advanced Search page.

Verify Location: To confirm that the probe set 1422223 targets the last exon of *Grin2b*, click on this link. The sequence on the probes on the array are immediately sent to the Genome Browser at the University of California, Santa Cruz, and the best match on the genome is found in real time using the BLAT algorithm. The BLAT Search Results page will open and if you want to drill down to a view of the genome, click on the "browser" link to the far left (top row). You will see a series of black boxes labeled "YourSeq" (your query sequence) and beneath that you will see several blue "tracks" labeled Grin2b and GRIN2B. If you zoom out 10X you will see that the probe set (YourSeq) actually targets the 3' end of the last exon--a bit more detail than we had before.

Basic Statistics: This button will generate summaries such as the average expression, the range, bar charts of expression ordered by strain and by rank.

Similar Traits: This button will provide you, in a separate window, a link to *Grin2b* expression data in other data sets that may interest you.

Probe Tool: A link to the sequence data for the individual probes that make up a probe set. This table can be used for a very fine-grained analysis of particular probe sets.

Update Data: A link to the database that allows modification of entries. You need to be logged in to GeneNetwork and to have a valid account to see this button.

Add to Collection: If you would like to add a trait to your collection of traits, transcripts, or markers, use this button.

Probe information

Most of the gene expression data available in GeneNetwork/WebQTL is derived from hybridization with Affymetrix microarrays. These arrays use one or more probe sets, each of which includes 11 to 16 25-base oligonucleotide probe sequences to represent each gene. Each probe set also includes, for each probe sequences, a mismatch probe sequence with one altered base as a control for nonspecific hybridization. Click the *Probe Tool* button for access to information about the probe sequences in this probe set.

The Probe Information window that open has a list of information about the probe in the probe set. Odd-numbered probes, shown in green, are the probes that match sequences in the target gene. Even-numbered probes are the mismatch control probes. This table provides a wide range of information about the properties of each probe and the data derived from it. Click on the title of each column for more information about information in the column.

Click the *Verify UCSC* button to open a window at the UCSC Genome Browser with the positions of the probe sequences marked in comparison to other features of the mouse genome.

Click the *Verify Genbank* or *Verify Ensembl* button to open a similar window at Genbank or Ensembl.

Click the *Select PM* button and then the *Correlation Matrix* button to display a correlation matrix of the expression estimates provided by individual probes. The correlation of individual probes may be poor. Among other things, the binding affinities of individual probes will differ, making them most sensitive in different ranges to changes in concentration of their target sequences.

Question: Why might two overlapping probes have highly negative correlations?

Basic statistics

Close the Probe Information window and click the *Basic Statistics* button.

This opens a Basic Statistics window with a table of trait values (mean values for each recombinant inbred line), a box plot summarizing the range of trait values across lines, two bar charts showing the strain means and standard errors (when available), and a probit or Q-Q plot that displays to what extent the trait distribution differs from normal. In all charts, the Y axis is a log scale, with 8 representing average expression for all genes, and each unit representing a two-fold change. Notice the range of expression among RI lines. Since the Y axis uses a log scale, the value 0 has no special meaning, and, in particular, does not define “unexpressed”. In fact, values below 5 often represent expression too low to be detected or estimated reliably.

The box plot is a simply summary of the spread of the values in the Trait Data and Analysis Form. The blue plus sign represents the mean expression, and the box defines the 25% and 75% quantiles. Click on the link beneath the plot for more information. The bar charts provide alternative

views of the same data. The error bars may be large, and are often based on two samples (in this case the SEM is usually the same as the SD). Also note that the size of these error bars tend to increase as the expression increases (non-uniform error). High noise and non-uniform variance are characteristics that suggest cautious interpretation.

Notice that some recombinant inbred lines have more extreme phenotypes than the parental lines (most easily seen in the second bar chart). This phenomenon is described by the slightly ominous term transgression. It suggests that the trait is controlled by several loci with opposing effects, loci which partly balance each other in the parental strains.

In the Normal Probability Plot, the ranked trait values are plotted on the ordinate against the expected values for each rank on the abscissa. If the data is normally distributed, the plotted points will approximate a straight line. If the plotted points form a non-symmetric curve, the data is skewed. If the plotted points are symmetric but appear to have a plateau in the center, the distribution is narrower than normal. If they are a symmetric S shape, the distribution is wider than normal. This last pattern is also characteristic of a trait that is strongly controlled by a single gene, because the distribution is actually a mixture of two distributions, one for each allele of the controlling gene.

Outliers in the data appear as points that are significantly above or below the line formed by the rest of the points.

Close the Basic Statistics and the Trait Data and Analysis Form window by clicking the close buttons. The should leave you with a Search Results page open; by clicking on the back button on this page (which will preserve your database choice), or by choosing *Search* > *Search Databases* from the menu (which will set the default database), you can return to the Find Records page.

Finding genetic loci that affect trait variation

Simple interval mapping

The function of the WebQTL section of GeneNetwork is QTL mapping. QTL mapping is, in essence, the search for correlation, among related individuals, between values of a trait and alleles of a marker locus somewhere in the genome. A high correlation suggests that there is a gene near the marker locus that affects the trait value. The estimated region in which that hypothetical gene might be found is a quantitative trait locus or QTL.

At the Find Records page, search for transcripts of the gene *Fprl1* in the BXD data set *INIA Brain mRNA M430 (Jan06) PDNN*. Click on the entry for probe set *1428589* to open a Trait Data and Analysis Form for that transcript.

Map Viewer

Find the *Interval Mapping* button and click it (you may have to scroll down a little to see it). After a short wait, a new Map Viewer window will open with a graph. The horizontal axis of this graph represents the mouse genome, and it is divided into separate sections for each chromosome. The heavy blue line represents the statistical significance (likelihood ratio statistic or LRS) of a hypothetical QTL at that location. Pink and gray horizontal lines show the thresholds at which the LRS is significant or suggestive, respectively. The thinner red or green lines represent the estimated strength of the hypothetical QTL.

For *Fprl1* you will see a strong QTL on chr 2. This is a so-called *trans* QTL because it is not near the location of the *Fprl1* gene, which is on chr 17. An blue triangle on the horizontal axis at chr 17 shows the position of the *Fprl1* gene.

At the top of the Map Viewer page, above the figure, is a box with several settings with which you can customize the figure.

Single-chromosome map

Click on the numeral *2* at the top of the chr 2 section of the graph. This will open a window with a Map Viewer for chromosome 2 and, below that, the Interval Analyst table for chr 2.

The Interval Analyst displays extensive additional information about genes and genetic variation on that chromosome. Genes in the region of the QTL peak (blue line) are candidate genes that may affect expression of the transcript being mapped.

In the single-chromosome map viewer, immediately above the figure, you will see pastel red, blue, and yellow lines spanning the width of the figure. Clicking on one of these lines will open another window corresponding to the region of the map around the click point. A click on the red line will open another Map Viewer at a higher scale. A click on the blue or yellow lines will display a new window with the corresponding region in the USCS or Ensembl genome browsers, respectively. These displays offer additional ways to examine genes that may affect the expression of the transcript you are mapping.

When your curiosity about this analysis is satisfied, close all windows except the Search Results and use the *Back* button to return to a Find Records page

Pair-scan mapping

Introduction

Complex traits often may be controlled by more than one locus, and, in fact, it is somewhat unrealistic to fit a single-locus to a trait and expect to get an accurate description of how it is controlled. A pair-scan is a step toward detecting greater complexity. It search for pairs of loci that can explain the trait variation. An exhaustive search, testing all available pairs of loci, would be time-consuming; WebQTL tests a grid of locations and then refines the search by testing on successively finer grids in areas that successfully explain trait variation.

Pairs of loci can affect a trait in two ways. They may contribute independent additive effects, so that the trait value can be determined by adding constant effects representing each allele at each locus. On the other hand, alleles at one locus may affect the effect of the other, so that the final trait value cannot be expressed as a simple sum of allele effects. This situation is described as epistasis or interaction.

The results of this mapping are displayed in a square graph or “heat map” that is divided by a diagonal line. The vertical axis represents the genomic location of one locus; the horizontal axis, the location of the other. Color in the figure represents the LRS for the association of the trait value with the pair of loci represented by that location. Warm colors, especially red, represent higher LRS values and therefore represent possible locations of a pair of interacting QTLs.

The area above and to the left of the diagonal line represents interaction effects alone. The area below and to the right represents the total effect of possible QTLs, including both additive and interaction

effects. A vertical bar to the right of the interaction map is the "heat map" equivalent of an interval mapping plot. It shows the results of searching for a single QTL to explain the trait.

Example

On the Find Records page, search for *Zc3hav1* (in the *BXD*, *Whole Brain mRNA*, *INIA Brain mRNA M430 (Jan06) PDNN* database). From the choices returned, click on *Probeset/1446244_at_A* to bring up the Trait Data and Analysis Form. Scroll down to the Pair Scan section. Choose *LRS Interact* in the *Sort by* menu. Click the *Pair-Scan* button.

After a period of computation, WebQTL will open a Pair-Scan Results window with the figure described above. For *Zc3hav1*, There is a strong QTL on chr 4 and a weak QTLs on chrs 16 and 19. These appear as lines of yellow and red in the lower-right triangle of the interaction plot. In addition, there are locations of a few possible interactions, chrs 1 and 2 with chr 19 and two locations on chr 6. These appear as spots of red in the rectangles representing those pairs of chromosomes, on both sides of the diagonal line. In the lower-right, the interaction appears as a red dot on the yellow line representing the weak chr 19 QTL. Although these interactions suffice for a demonstration, they are not robust and may not be reproducible.

Scroll down to the table below the figure. This table lists pairs of loci in descending order of interest. If you specified that the list be ordered by interaction effect (by choosing *Sort by LRS Interact*), the top two entries will be the chr 1 - chr 19 and chr 2 - chr 19 locus pairs.

Scroll up to the figure again and click anywhere in the rectangle representing chrs 2 and 19 in the upper-left triangle. WebQTL will open another window with an expanded map of just that chromosome pair. This map is created without the short-cut sampling used for the whole-genome pair-scan. That is, all pairs of available markers for chrs 2 and 19 are evaluated to produce this map.

To find genetic correlations among traits

If a trait is measured in a set of recombinant inbred lines, differences among the strains can be attributed, in part, to their genetic differences. Environmental and stochastic differences can be minimized by testing under constant environmental conditions and by averaging trait values across several individuals. Under these conditions, traits that are affected by the same genes, directly or indirectly, would show similar variation across different strains because they would be responding to the same allelic differences in the controlling genes. Thus, traits that are affected by similar sets of genes would be expected to be correlated across a set of recombinant inbred lines. These traits would be expected to be functionally related in some way. GeneNetwork allows search for traits that may be functionally related by searching for those whose average trait values are correlated.

Return to the a Find Records page (choose *Search* > *Search Databases*) from the GeneNetwork menu and search for *Csf2ra* in the *BXD*, *Whole Brain mRNA*, *INIA Brain mRNA M430 (Jan06) PDNN* database.

Choose *1420703_at_A*, the probe set that targets *Csf2ra* on chr 13. The probe set that targets it on chr 2 seems to be nonspecific. In the Trait Data and Analysis Form, scroll down to the Trait Correlations section. Under *Choose Database*, choose *BXD Published Phenotypes*. Click the *Trait Correlations* button.

The Correlation Results page that opens shows a table listing published traits for the BXD RI set and their correlations with the *Csf2ra* transcript. By default, the traits are listed in order of descending

P-value. This transcript shows a relatively high negative correlation with brain weight for one data set. In fact, this example was chosen for this correlation. It also shows a high positive correlation to an alcohol-related trait. Farther down the list, it also shows lesser negative correlations with various measures related to brain size in other data sets.

Click on the value for the *Csf2ra*-brain weight correlation (*-0.59*). This action will open a Correlation Plot page in which you can examine the relationship between the two traits. Look for linearity and outliers.

Selection and saving multiple traits

The list of traits on the Correlation Results page represents traits that may be related in some way. You may want to select a group of them for further analysis. For example, Use the checkboxes to the left of each entry to check brain-related entries 1, 4, 8, 10, 13, 15, 17. Click the *Add to Collection* button at the top of the page. This button will add the checked traits to your BXD Trait Collection page.

Multiple QTL mapping

Close the BXD Trait Collection Page and return to the Correlation Results page. This page also provides direct access to two multiple-trait analysis functions. With seven traits still checked on this page, click the *Multiple Mapping* button at the top of the page. This button will open a Multiple Interval Mapping page that shows QTL scans for all selected traits on the same figure. In this example, a few traits seem to share weak QTLs on chr 6 and possibly also chr 8.

QTL Cluster Map

Close the Multiple Interval Mapping window and return once again to the Correlation Results page. With seven traits still checked, click the *QTL Cluster Map* button at the top of the page. This button opens a QTL Cluster Map page with a figure that combines trait clustering and QTL mapping functions. Traits are clustered according to their pairwise correlation and a QTL scan is performed for each trait. The results of the scan are displayed as a vertical bar where bright color indicates the location of potential QTLs. The clustering places the QTL maps for related traits closer to each other. The arrangement allows you to recognize control regions that would not be individually significant but which become noteworthy if they appear in many related traits.

Correlation of expression among genes

Close the QTL Cluster Map and Correlation Table pages and return to the Trait Data and Analysis page. Choose *Search* > *Search Databases* from the GeneNetwork menu.

Search for *Lin7c* (in the *BXD*, *Whole Brain mRNA*, *INIA Brain mRNA M430 (Jan06) PDNN* database) and choose probe set *1450937* from the search results. In the Trait Data and Analysis Form, the default database should be *INIA Brain mRNA M430 (Jan06) PDNN*. Click the *Trait Correlations* button. This search will take a little longer because it is searching a large gene expression data set. It will return a list of 100 genes that all show high correlation, positive or negative, with *Lin7c*. Click the *Select All* button.

WebGestalt

WebGestalt is a Web-based gene set analysis toolkit. GeneNetwork provides an easy way to submit to WebGestalt a set of genes related by correlated expression. We will explore only one WebGestalt function.

With 100 expression traits selected, click the *WebGestalt* button at the top of the page. This action will open a page from WebGestalt that redisplays the input data and displays links to all the WebGestalt analysis functions. When this page displays, notice the section in the center of the page entitled *Gene set analysis tool*. The *GO Tree* button provides an analysis of a gene set in terms of the gene ontology categories for the genes in the set. Click the *GO Tree* button. When the process appears to be complete, click the *Check GO Tree* button.

A gene ontology is a hierarchical categorization of genes by their functions. A large subset of the roughly 20000 genes measured using microarrays have been assigned to one or more functional categories. The three independent dimensions of this ontology are "biological process", "molecular function", and "cellular component". WebGestalt will analyze the gene ontology categories and display a hierarchical list of categories. Categories in which genes of the submitted set appear preferentially will appear in red. These functional categories characterize genes whose expression is correlated with *Lin7c*. Categories can to opened by clicking on them to display information about more specific sub-categories. For example, open the biological process > physiological process > localization category hierarchy.

Analysis functions for multiple traits

Open your BXD Trait Collection window by choosing *Search* > *Trait Collections* > *BXD Collection* from the GeneNetwork menu. (If it does not appear, it may already be open beneath other windows.) If there are trait entries in the window, remove them by clicking the *Select All* and *Remove Selection* buttons.

Return to a Find Records page or open one with *Search* > *Search Databases*. Search for *App* in the *BXD*, *Whole Brain mRNA*, *INIA Brain mRNA M430 (Jan06) PDNN* database. Choose *1420621_a_at_A* from the Search Results page.

In the Trait Data and Analysis Form, scroll down to the Trait Correlations section and choose *BXD Published Phenotypes* for the database. Click the *Trait Correlations* button.

The Correlation Results page that opens provides you with a number of classical traits that are correlated with differences in the *App* gene. Choose five to ten of these by checking the checkboxes at the left of each. Click the *Add to Collection* button at the top of the page.

Having found a group of classical traits correlated with *App*, we will now do the same for a group of gene expression traits. Return to the Trait Data and Analysis Form (close the Correlation Results page, if you wish). In the Trait Correlations section, change the database to *INIA Brain mRNA M430 (Jan06) PDNN*. Click the *Trait Correlations* button. This search will take longer.

The Correlation Results page that opens displays genes whose expression is correlated with that of *App*. Some of the correlations for this example are quite high. Choose five to ten of the highest-correlated genes by checking them and add them to the BXD Trait Collection page by clicking the *Add to Collection* button.

The BXD Trait Collection page now has both classical and gene expression traits all of which correlate with the expression of *App*. We would expect some of these to be functionally related. We can now explore some GeneNetwork/WebQTL functions that help analyze such a group of potentially related traits.

Multiple QTL mapping

Choose about eight of the traits on the BXD Trait Collection page and click the *Multiple Mapping* button. The function performs a QTL scan for all selected traits and plots the result in the same figure. In the Multiple Interval Mapping page that opens, the different traits will be represented by color-coded lines, each of which plots the LRS for one trait. Look for regions of the genome where several of the lines have coincident peaks; these may represent the location of a control gene common to the mapped traits.

Correlation comparison

Return to the BXD Trait Collection page. Change the selection of traits if you want, and click the *Compare Correlates* button. A Correlation Comparison window opens with introductory explanation and opportunity to change options for the analysis. Using the default options, click the *Correlate* button in the middle of the page.

When the calculation finishes, the Correlation Comparision page redraws with two lists of results. These list groups of genes from the database whose expression is correlated with one or more traits in the submitted set. The potentially interesting groups are those in which a several database genes are correlated with several of the input genes. This feature can also be used to identify the genes that have common relations to a set of physiological or behavioral traits.

Correlation Matrix

Return to the BXD Trait Collection page. Change the selection of traits if you want, and click the *Correlation Matrix* button.

The Correlation Matrix page that opens shows a simple table with all pairwise correlation coefficients among the submitted traits. Values are color coded to help identify the more important correlations. The table cells also show the number of value pairs on which each coefficient is based, those coefficients based on few values may be unreliable.

This page also presents principal components calculated from the correlated traits if the number of data points for each trait is sufficient. If no principal components were generated, examine the table to identify the traits with fewest values. Close the Correlation Matrix window, return to the BXD Trait Collection page, uncheck the traits with few values, and click the *Correlation Matrix* button again. Principal components can be considered to be synthetic traits that summarize the common components of a group of correlated traits. Principal component traits can be transferred to the BXD Trait Collection page and used for further QTL mapping or correlation analysis.

Association Network

Return to the BXD Trait Collection page. Change the selection of traits if you want, and click the *Network Graph* button. This function creates an Association Network page with a graphical representation of the pairwise correlations among the submitted traits. In the graph, nodes represent

classical or gene expression traits, and lines connecting the nodes represent correlations. Lines are color-coded to indicate the sign and strength of the correlations.

This concludes the GeneNetwork tutorial. The following section summarizes the functions available under the GeneNetwork menu items.

GeneNetwork menu

Home

GeneNetwork Intro

This page provides a brief introduction to GeneNetwork/WebQTL. It describes how users can submit their own trait data for analysis.

Enter Trait Data

This page allows users to enter their own trait data for analysis, either by entering data into a form or by uploading a file.

Batch Submission

This page allows users to upload a file containing values for multiple traits. These are stored temporarily on GeneNetwork/WebQTL servers and are available for analysis during that time.

Search

Search Databases

This is the default starting page (rather than Home), and this is the page from which most users will begin using GeneNetwork/WebQTL. In this tutorial it is also called the Find Records page.

SNP Browser

This browser allows you to identify locations of single nucleotide polymorphisms among inbred mouse lines and genotypes of each line at those locations. This browser combines SNPs from Celera Genomics, the Perlegen/NIEHS resequencing project, the Wellcome-CTC SNP Project, dnSNP, and the Mouse Phenome Database.

GeneWiki

GeneWiki is a tool that allows users to annotate a gene or transcript with information that may be generally useful. Like information in any wiki, each annotation page is a public record that can be expanded or edited by any user. GeneWiki pages also display GeneRIF annotation from NCBI (which is not editable).

Interval Analyst

The Interval Analyst is a table listing, for a defined chromosomal interval, information about known mouse genes and variation in those genes.

GenomeGraph

The GenomeGraph tool displays figures that give a genome-wide picture of the control of gene expression. Each of these figures plots the location of a gene transcript against the location of the strong quantitative trait locus (QTL) that affects the expression of that transcript. The data for these

figures is precomputed from microarray data that estimates gene expression for many transcripts for each member of a genetic reference population.

Trait Collections

This submenu provides access to the lists of traits selected from searches. Each reference population has a separate collections page, which can combine classical traits, gene expression traits, and genotypes.

Scriptable Interface

The scriptable interface provides a programmer's interface to allow users to write scripts that automatically retrieve information or perform simple analyses with GeneNetwork/WebQTL.

Simple Query Interface

This page provides a Web form that is the equivalent of the scriptable interface for information retrieval.

Database Information

This submenu provides access to extensive descriptions of the data sets provided by GeneNetwork/WebQTL

Help

WebQTL Movie

A tutorial video, soon to be updated.

WebQTL Demonstration

A pair of Powerpoint presentations describing the use of GeneNetwork/WebQTL. One describes using GeneNetwork/WebQTL to explore trait variation and covariation, and the other describes QTL mapping in WebQTL.

WebQTL Tour

A Web tutorial focusing on QTL mapping in WebQTL. It is still useful, although GeneNetwork/WebQTL has evolved considerably since it was written.

WebQTL FAQ

A useful list of frequently asked questions and answers.

Glossary of Terms

Definitions and discussion of some critical terms used in GeneNetwork/WebQTL.

News

Announcements of new data sets or analytical functions in GeneNetwork/WebQTL.

References

Publications that describe GeneNetwork/WebQTL and its components and data sets, and publications that cite or use GeneNetwork/WebQTL. If you publish work based on GeneNetwork/WebQTL, check this list for appropriate citations.

Policies

Conditions and Limitations

Conditions and limitations on the use of data from GeneNetwork/WebQTL.

Data Sharing Policy

Links to guidelines and recommendations about sharing data obtained as part of large scale biological research projects.

Status and Contacts

A description of the status of various data sets that are part of GeneNetwork/WebQTL and contact information for the scientists who are responsible for their existence.

Privacy Policy

A description of small amount of information that is recorded about your use of GeneNetwork/WebQTL. Data submitted through routine use of GeneNetwork/WebQTL is never stored permanently.

Accounts

A few functions or specific data sets may be temporarily restricted to users who have been given password-protected accounts on GeneNetwork/WebQTL. If you contribute a data set to GeneNetwork/WebQTL, you can choose to have it restricted to selected users for a period of time (for example, until publication of a description of the data set).

Links

Links to a variety of Web-based genetics and genomics resources.

Acknowledgements

This work was supported by a Human Brain Project funded jointly by the National Institute on Drug Abuse, National Institute of Mental Health, the National Institute on Alcohol Abuse and Alcoholism, and the National Science Foundation (award P20-MH 62009 and P20-DA 21131 to KF Manly and R Williams) and by a separate grant from The National Institute on Alcohol Abuse and Alcoholism (INIA grants U01AA13499, U24AA13513 to R Williams).